HSR Health Services Research

DOI: 10.1111/1475-6773.12519
RESEARCH ARTICLE

Hospital Quality Reporting in the United States: Does Report Card Design and Incorporation of Patient Narrative Comments Affect Hospital Choice?

Martin Emmert and Mark Schlesinger

Objective. To explore the impact of hospital report card design and incorporation of patient narrative comments on consumers' choices of hospitals.

Data Sources. Primary data collected from an online survey with 1,350 respondents in February, 2015.

Study Design. A randomized 2 (narrative comments: yes, no) × 3 (design: representation of clinical performance in textual, star, numerical formats) between-subject online-based cross-sectional experiment.

Principal Findings. In 51 percent of all cases, respondents selected the hospital with the best clinical results. Report cards with a numerical design induced choices more focused on clinical ratings (56.0 percent chose the highest rated hospital) than those with textual information (48.1 percent) or star ratings (47.3 percent) ($p < .001$). Report cards without narrative comments (49.7 percent) and with narratives (51.4 percent) were not associated with significant difference in selecting top-rated clinical hospitals ($p = .376$). But there were significant interactions affecting choice of hospitals among exposure to narratives, formatting of clinical performance, and respondents' education.

Conclusions. Consumers have a difficult time synthesizing quality data in various formats. Hospital report cards continue to pose challenging choices, especially for those with limited education. Narrative comments in their earliest emerging forms do not seem to be altering hospital choice as much as the literature has suggested for other providers, but they may have consequential impact on the choices of certain subsets of consumers.

Key Words. Survey research and questionnaire design, medical decision-making, hospitals, quality improvement/report cards

Over the last two decades, much effort has been put into implementing and refining online report cards to create more transparency for health care quality (Lagu et al. 2010; Damberg and McNamara 2014). Yet patients have been

slow to take advantage of quality report cards in making choices among health care providers (Center for Advancing Health 2009; Hibbard, Greene, and Daniel 2010; Fox 2011). Given this checkered track record, the challenge seems to be to provide not only the information that patients want to have (Hibbard, Greene, and Daniel 2010) or as much information as possible (Hibbard and Peters 2003; Peters et al. 2007), but to understand how to present and target the information to enhance consumer engagement. Despite some key insights from prior research (e.g., Hibbard and Peters 2003; Hibbard, Greene, and Daniel 2010; Mehrotra et al. 2012), we still do not fully understand how best to design public report cards (Damberg and McNamara 2014; Schlesinger et al. 2014).

An additional complication has arisen in recent years: the spread of anecdotal consumer comments about health care providers, available over the Internet on a variety of consumer websites (Schlesinger et al. 2015). In these accounts, patients write open-ended descriptions of their experience with providers, expressed in their own words. Incorporating narrative comments from patients has been suggested as one possible strategy for making quality-reporting websites more accessible to consumers who are less numerate or sophisticated in their decision-making (Hibbard and Peters 2003; Lagu and Lindenauer 2010; Lagu et al. 2013; Shaffer, Owens, and Zikmund-Fisher 2013; Greaves, Millett, and Nuki 2014). Such comments are also expected to provide a more complete picture of the total patient experience with that provider, incorporating emotional reactions and the meaning that patients ascribe to their experiences (Brennan 1995; Greaves, Millett, and Nuki 2014; Schlesinger et al. 2015).

Despite this promise (Lagu and Lindenauer 2010; Martino 2012; Greaves, Millett, and Nuki 2014), we know relatively little regarding the importance of narrative comments as well as their actual impact on patient perceptions, understanding, or behavior. The limited literature suggests that: (1) it is challenging for many consumers to integrate diverse measures and representations of quality, when selecting among health care providers (Hanoch et al. 2009; Brutus 2010; Schlesinger et al. 2014) and (2) at least under some circumstances, narratives appear to affect consumer choices (Winterbottom

Address correspondence to Martin Emmert, Ph.D., Department of Health Policy and Management, Yale University School of Public Health; 47 College Street, New Haven, CT 06520; e-mails: martin.emmert@yale.edu; martin.emmert@fau.de. Martin Emmert, Ph.D., is also with the School of Business and Economics, Institute of Management (IFM), Friedrich-Alexander-University Erlangen-Nuremberg, Nuremberg, Germany. Mark Schlesinger, Ph.D., is with the Department of Health Policy and Management, Yale University School of Public Health, New Haven, CT.

et al. 2008; Shaffer, Owens, and Zikmund-Fisher 2013; Greaves, Millett, and Nuki 2014), especially among certain subsets of the public (Greaves, Millett, and Nuki 2014). But very little is understood about the impact of narratives in particular health care settings or how that impact varies with the scope and format of other performance metrics.

Study Hypotheses

From this limited extant literature, we derive three hypotheses regarding the likely impact of patient narratives on the choice among hospitals. First, the presence of narratives is likely to alter the attention consumers pay to other performance metrics (Kanouse et al. 2016). But this influence could either enhance or undermine attention to standardized measures. Narratives are inherently more variable in their valence (i.e., positive vs. negative assessment), because they present experience at the individual level, whereas standardized metrics present averages or frequency distributions. Consumers might gravitate to the more nuanced information in narratives, becoming too cognitively overloaded to pay much attention to quantified ratings. On the other hand, consumers might find the narratives themselves to be too complex, turning instead to standardized metrics because they simplify comparisons. Because it is unclear which effect will dominate, we hypothesize that there will be an effect, but not the direction that impact will have.

Our second hypothesis relates to the format in which standardized metrics are presented. Prior to the emergence of narratives, symbolic representation (e.g., star ratings) of standardized metrics was commonly favored to help consumers process multiple performance measures (Hibbard and Sofaer 2010; Hibbard et al. 2012). But narratives involve text; it may therefore create more cognitive dissonance for consumers to switch between narrative and symbolic representations of quality (Brutus 2010; Damman et al. 2011). Consequently, we hypothesize that the impact of including narratives on use of other measures (e.g., clinical quality measures, patient satisfaction, or safety metrics) will be most pronounced when those measures are presented in numeric or symbolic format, rather than as text.

Finally, we anticipate that narratives will have the largest impact on consumers who are less educated—who might otherwise be overwhelmed by numeric or symbolic quality metrics. These sorts of effects have been documented for the inclusion of narratives in decision aides related to treatment choice (Winterbottom et al. 2008; Dieckmann, Slovic, and Peters 2009). We hypothesize that this will extrapolate to choices among providers, based on

evidence that consumers' ability to comprehend comparative performance metrics affects provider choice (Schlesinger et al. 2014).

METHODS

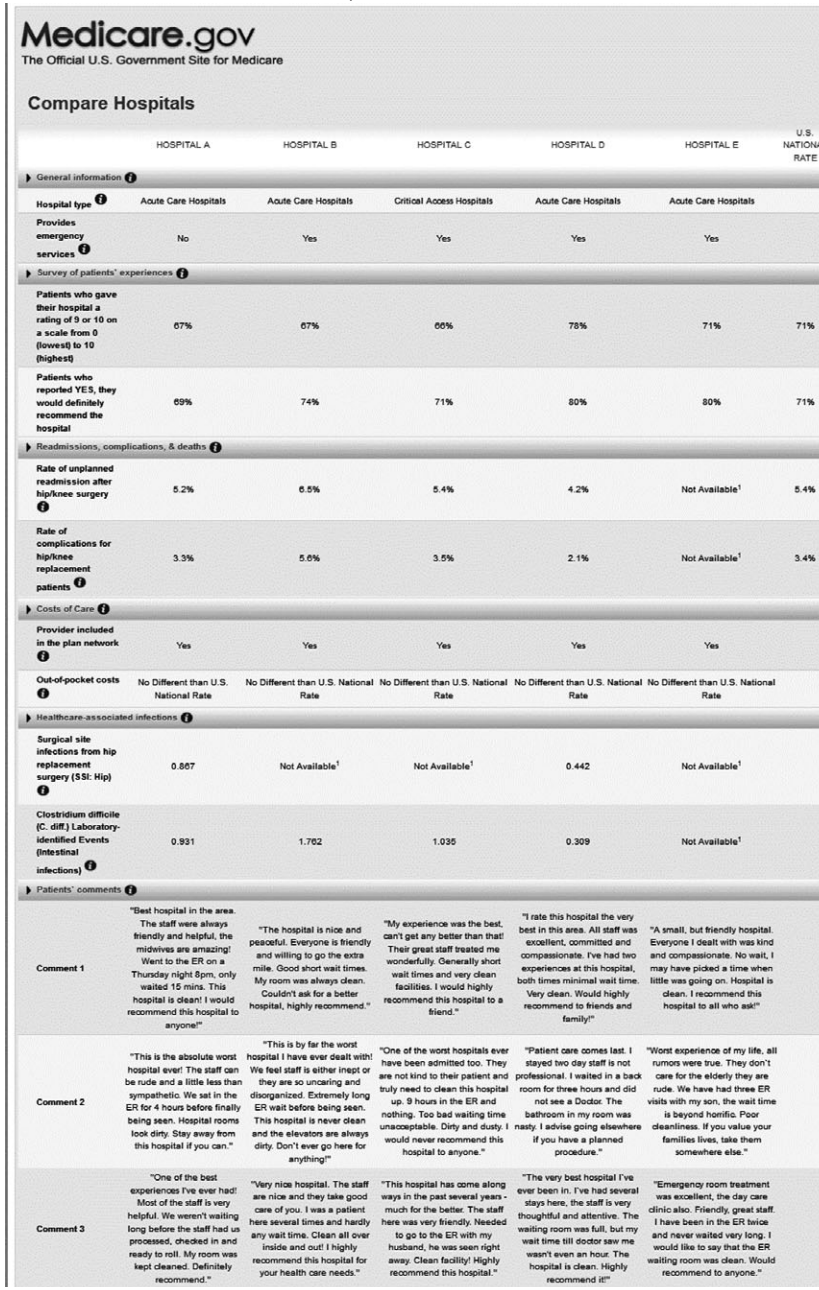
This study was designed as a 2 (narrative comments: yes, no) × 3 (design: clinical performance metrics in textual, star, numerical formats) between-subjects online-based experiment wherein participants were randomly assigned to two of the six conditions. Each report card described five hospitals and contained one dominating as well as one dominated hospital in terms of clinical quality of care, the dominating hospital performing equal or better on all dimensions as the other four hospitals in the choice set. After reviewing each report card, respondents were asked to choose among the five hospitals, weighing the importance of multiple hospital attributes and quality metrics, and (in the experimental arms that incorporated comments) reviewing patient narratives. The main outcome measure was the selection of the best (dominating) or worst (dominated) performing hospital, in terms of their clinical metrics.

To pose choices in a realistic fashion, we followed Lagu et al. (2013) and modified Hospital Compare, the official U.S. government site for Medicare. Within this template, we created a decision framework that was reasonably simple—to avoid overwhelming respondents with information (Slovic 1982)—but nonetheless incorporated most of the dimensions of hospital performance demonstrated in past research to be relevant to consumers. This include *structural information* (attributes of the hospital from which consumers might infer quality), metrics of *technical quality of care* (HEDIS-like measures), *patient satisfaction* (HCAHPS-like measures), *costs of care* and *safety metrics* (Figure 1). To assign realistic information, we derived data from the Hospital Compare website, though assigned it to unnamed hypothetical institutions.

Representative Narrative Comments

We created a standardized set of narrative comments to incorporate into the report cards in three of the six experimental arms. These were based on (de-identified) real-world comments about U.S. hospitals, selected to be typical in terms of properties demonstrated to affect interpretation of health care–related narratives (e.g., consistency, valence, length, and complexity) (Winterbottom et al. 2008; Archak, Ghose, and Ipeirotis 2011; Detz, López, and

Figure 1: Screenshot of Hospital Report Card: Design 4 (Numerical information with narrative comments)



Medicare.gov

The Official U.S. Government Site for Medicare

Compare Hospitals

	HOSPITAL A	HOSPITAL B	HOSPITAL C	HOSPITAL D	HOSPITAL E	U.S. NATIONAL RATE
General information						
Hospital type	Acute Care Hospitals	Acute Care Hospitals	Critical Access Hospitals	Acute Care Hospitals	Acute Care Hospitals	
Provides emergency services	No	Yes	Yes	Yes	Yes	
Survey of patients' experiences						
Patients who gave their hospital a rating of 9 or 10 on a scale from 0 (lowest) to 10 (highest)	67%	67%	66%	78%	71%	71%
Patients who reported YES, they would definitely recommend the hospital	69%	74%	71%	80%	80%	71%
Readmissions, complications, & deaths						
Rate of unplanned readmission after hip/knee surgery	5.2%	6.5%	5.4%	4.2%	Not Available[1]	5.4%
Rate of complications for hip/knee replacement patients	3.3%	5.6%	3.5%	2.1%	Not Available[1]	3.4%
Costs of Care						
Provider included in the plan network	Yes	Yes	Yes	Yes	Yes	
Out-of-pocket costs	No Different than U.S. National Rate	No Different than U.S. National Rate	No Different than U.S. National Rate	No Different than U.S. National Rate	No Different than U.S. National Rate	
Healthcare-associated infections						
Surgical site infections from hip replacement surgery (SSI: Hip)	0.867	Not Available[1]	Not Available[1]	0.442	Not Available[1]	
Clostridium difficile (C. diff.) Laboratory-identified Events (Intestinal infections)	0.931	1.762	1.035	0.309	Not Available[1]	
Patients' comments						
Comment 1	"Best hospital in the area. The staff were always friendly and helpful, the midwives are amazing! Went to the ER on a Thursday night 8pm, only waited 15 mins. This hospital is clean! I would recommend this hospital to anyone!"	"The hospital is nice and peaceful. Everyone is friendly and willing to go the extra mile. Good short wait times. My room was always clean. Couldn't ask for a better hospital, highly recommend."	"My experience was the best, can't get any better than that! Their great staff treated me wonderfully. Generally short wait times and very clean facilities. I would highly recommend this hospital to a friend."	"I rate this hospital the very best in this area. All staff was excellent, committed and compassionate. I've had two experiences at this hospital, both times minimal wait time. Very clean. Would highly recommend to friends and family!"	"A small, but friendly hospital. Everyone I dealt with was kind and compassionate. No wait, I may have picked a time when little was going on. Hospital is clean. I recommend this hospital to all who ask!"	
Comment 2	"This is the absolute worst hospital ever! The staff can be rude and a little less than sympathetic. We sat in the ER for 4 hours before finally being seen. Hospital rooms look dirty. Stay away from this hospital if you can."	"This is by far the worst hospital I have ever dealt with! We feel staff is either inept or they are so uncaring and disorganized. Extremely long ER wait before being seen. This hospital is never clean and the elevators are always dirty. Don't ever go here for anything!"	"One of the worst hospitals ever have been admitted too. They are not kind to their patient and truly need to clean this hospital up. 9 hours in the ER and nothing. Too bad waiting time unacceptable. Dirty and dusty. I would never recommend this hospital to anyone."	"Patient care comes last. I stayed two day staff is not professional. I waited in a back room for three hours and did not see a Doctor. The bathroom in my room was nasty. I advise going elsewhere if you have a planned procedure."	"Worst experience of my life, all rumors were true. They don't care for the elderly they are rude. We have had three ER visits with my son, the wait time is beyond horrific. Poor cleanliness. If you value your families lives, take them somewhere else."	
Comment 3	"One of the best experiences I've ever had! Most of the staff is very helpful. We weren't waiting long before the staff had us processed, checked in and ready to roll. My room was kept cleaned. Definitely recommend."	"Very nice hospital. The staff are nice and they take good care of you. I was a patient here several times and hardly any wait time. Clean all over inside and out! I highly recommend this hospital for your health care needs."	"This hospital has come along ways in the past several years - much for the better. The staff here was very friendly. Needed to go to the ER with my husband, he was seen right away. Clean facility! Highly recommend this hospital."	"The very best hospital I've ever been in. I've had several stays here, the staff is very thoughtful and attentive. The waiting room was full, but my wait time till doctor saw me wasn't even an hour. The hospital is clean. Highly recommend it!"	"Emergency room treatment was excellent, the day care clinic also. Friendly, great staff. I have been in the ER twice and never waited very long. I would like to say that the ER waiting room was clean. Would recommend to anyone."	

Sarkar 2013; King, Racherla, and Bush 2014; Kanouse et al. 2016). To ensure that the substance of comments presented on the website would feel realistic to website users, we categorized a random sample of 1,000 narrative comments about hospitals from RateMDs.com into the five most frequently mentioned topics (i.e., general impression of the hospital stay, demeanor of clinicians and staff, wait time within hospital, facility cleanliness, and patients' recommendations) and selected the set of comments presented on the website to be representative of the naturally occurring prevalence of each of these topics.

For each hospital, we displayed two positive and one negative narrative comment—since positive comments appear more frequently in health care provider–related narratives (Lagu et al. 2010, 2013; Emmert et al. 2014a; Schlesinger et al. 2015). Because the distribution of valences was constant across all five hospitals, it was uncorrelated with their quantified quality metrics. In real life, the valence of patient comments (mix of positive and negative comments) tends to have a modest positive correlation with other quality indicators, but given the small sample displayed in this report card, setting the correlation to zero was a reasonable approximation.

Because the modal valence of narratives was the same for all hospitals, the comments presented here offered a discordant "signal" of quality compared to the standardized quality metrics. Comments in this design did not single out a preferred alternative to the hospital with the best star-ratings but could plausibly reduce the value consumers place on the ratings by suggesting that alternative quality orderings were plausible. As narratives are first introduced to quality reporting and the number of comments for any one provider remains small, this added fuzziness to quality rankings seems the most likely influence of narratives on consumer choice.

Survey Instrument

We designed the survey by using Questback's Internet–based EFS Survey software. The questionnaire consisted of four parts. After collecting some socio-demographic information, we asked about respondents' (aged 18 or older) past hospital search behavior and experience with hospital report cards. Respondents then completed the choice task, after which we assessed their acceptance and use of hospital report cards for future decision-making (not presented here). Before conducting the study, the questionnaire was piloted by 50 individuals and refined the survey in response to this feedback. The

survey was administered and conducted by Soapbox Sample, a fieldwork agency for survey research. The surveyed panel members are recruited from multiple channels (e.g., online, telephone, mobile, face-to-face, social media, print) and its demographics are similar to the U.S. population (U.S. Census 2012).

Data Analysis

Results are presented as both mean and standard deviation for parametric data and as numbers and percentages for nonparametric data. We performed comparisons between more than two groups by applying the Kruskal–Wallis test for non-normally distributed data. (The Shapiro–Wilk test was applied to examine the normality of the data distribution.) Additionally, we used chi-square for nonparametric data (two-sided).

Because the primary comparisons across report card designs and inclusion of narrative comments involved randomly assigned experimental exposures, our primary results involve simple comparisons across experimental arms. However, to identify potential underlying sources for observed cross-arm differences in choice, we also estimated a set of multivariate logistic regression models. The primary analyses focused on the probability that participants selected the quantitatively dominant or dominated hospital.

To facilitate the exploration of these multivariate findings, we estimated a sequence of models, each stage introducing an additional set of variables demonstrated by past research to affect how consumers address complex health-related decisions: (1) baseline model (including only indicator variables for experimental arms); (2) adding measures of demographics (age, gender, marital status, education, ethnicity, health insurance) that shape individuals' experience with and attention to health-related matters—and therefore their capacity to make sense of a complex choice context (Faber et al. 2009; Hibbard et al. 2012; Damberg and McNamara 2014); (3) adding health-related experiences (chronic conditions and number of hospital treatments in the previous 3 years) that could also enrich consumers' capacity to interpret hospital quality indicators (Schlesinger et al. 2012); (4) adding decision-making style (low vs. high maximizers; perceptions of quality differences among hospitals) that have been shown to influence how consumers incorporate multiple quality metrics into health care choices (Schlesinger et al. 2014). All statistical analyses were conducted using *SPSS version 22.0* (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Armonk, NY, USA). Observed differences were

identified as statistically significant if $p < .05$, and highly significant if $p < .001$.

RESULTS

Our final study sample consisted of 1,350 respondents who completed the survey experiment (completion rate = 65.57 percent). Thirty-four participants were excluded from subsequent analysis because of an extremely short answer time and/or inconsistent answer patterns. The overall mean age was 45.39 (±16.90) years, 724 respondents were female (53.6 percent), 675 were married (50.0 percent), and 211 respondents had a less than high school educational level (15.6 percent) (see Table 1). Our surveyed sample matched the U.S. Population (US Census 2012) in terms of age, gender, education, and marital status. The online surveys averaged 16.35 (±10.19) minutes.

The Impact of the Report Card Design on the Choice Behavior

Respondents selected the hospital with the best clinical results in slightly more than half of all cases (50.6 percent; 1,365 of 2,700). Significant differences were evident among the three report card designs ($p < .001$); those with a numerical formatting yielded more reliable choices of the quantitatively dominant hospital (56.0 percent) than those with textual information (48.1 percent) or a star rating display (47.3 percent) (see Table 2). By contrast, respondents selected the hospital with dominated quantitative metrics in approximately 9 percent of all experimental choices, with modest variation across report card design (overall 8.6 percent: numerical design: 10.1 percent, star display: 8.5 percent, textual information: 7.1 percent; $p = .072$).

Significant variation in choices was evident among different sociodemographic groups. For example, female respondents, those with higher education levels, private health insurance coverage, and those without any chronic conditions were more successful in selecting the hospital with the highest clinical ratings (see Table S1). The educational gradient—to which we return below—is largest for selection of the dominant hospital (35 percent for those not completing high school; 60 percent for those who had completed college) and is most pronounced at the lower end of the educational distribution.

Table 1: Overview of the Study Sample (p Value Was Calculated Using Chi-Square Test)

Characteristics	*Study Sample (N = 1,350)*	*U.S. Population*[†]	*p*
Age			
18–24 years	13.4%	13.1%	.807
25–34 years	18.1%	17.6%	
35–44 years	16.6%	16.9%	
45–54 years	18.6%	18.6%	
55–64 years	17.5%	16.1%	
65 years and older	15.9%	17.7%	
Gender			
Male	46.4%	48.9%	.179
Female	53.6%	51.1%	
Marital status			
Married	50.0%	50.7%	.957
Widowed	6.1%	5.7%	
Divorced	10.7%	10.1%	
Separated	2.1%	2.3%	
Never married	31.1%	31.2%	
Educational attainment			
Less than high school	15.6%	17.6%	.215
High school graduate	31.6%	28.5%	
Some college or associate's degree	26.7%	27.2%	
Completed college/advanced degree (bachelors, masters, professional, doctorate)	26.1%	26.8%	
Race/ethnicity			
White	80.5%	76.4%	<.001
Black or African American	9.2%	13.6%	
American Indian and Alaska Native	1.2%	1.6%	
Asian	5.5%	5.6%	
Some other race	3.6%	5.7%	
Household size			
One-person household	20.1%	26.7%	<.001
Two persons	32.7%	32.8%	
Three persons	20.5%	16.1%	
Four persons	15.5%	13.4%	
Five or mor persons	11.2%	11.0%	
Health insurance coverage			
Private plan (direct-purchase)	9.7%	11.0%[‡]	
Private plan (employment-based)	39.0%	53.9%[‡]	
Medicare	24.7%	15.6%[‡]	
Medicaid	11.6%	17.3%[‡]	
Military health care	1.9%	4.5%[‡]	
Uninsured	13.1%	13.4%	

Continued

Table 1: *Continued*

Characteristics	*Study Sample (N = 1,350)*	*U.S. Population*[†]	*p*
Chronic conditions			
Heart disease (ever)	7.1%	6.1%/10.8%[§]	
Asthma (current)	12.7%	12.7%	
Diabetes (ever)	13.7%	8.6%	
Arthritis-related conditions (ever)	19.0%	20.6%	
Hypertension (2 + visits)	20.4%	23.9%	
High cholesterol (ever)	27.6%	n.a.	
Any chronic condition	54.1%	n.a.	
Frequency of medical treatment in a hospital (past 3 years)			
No treatment in the past 3 years	36.8%	n.a.[¶]	
Once	22.9%	n.a.[¶]	
2–3 times	24.6%	n.a.[¶]	
4–6 times	7.9%	n.a.[¶]	
7–10 times	3.4%	n.a.[¶]	
11 times or more	2.2%	n.a.[¶]	
I wish not to answer this question	2.1%	n.a.	

[†]Age, Gender, Marital status, Education, Health Insurance Coverage, Ethnicity are derived from the U.S. Population—U.S. Census Bureau—Data 2012; Current population survey, Annual Social and Economic Supplement, 2012; American Community Survey (ACS) 2012; Chronic conditions according to the CDC National Health Interview Survey (2012).
[‡]The percentages do not sum up to 100% since roughly one-fifth of the population (18.1%) had multiple coverage types during the year.
[§]The lower value includes coronary heart disease, angina, or heart attack; the higher value includes coronary heart disease, angina, heart attack, or any other heart condition or disease.
[¶]The CDC National Health Interview Survey (2012) refers to hospitals stays in the past year.

Preferences and Choices Across Experimental Arms

Respondents' reports on the factors they viewed as more relevant to their choices offers some potentially useful diagnostic information regarding the emergence of these differences in choice across report card design. Each data element in the report card was scored on a 1–5 scale (1 = not all important; 5 = extremely important). Across arms, respondents reported that health care associated infections rates (Surgical site infection: 4.12 (±0.99), Intestinal infections: 4.09 (±1.01)) as well as technical quality of care information (Rate of unplanned readmission: 3.98 (±1.00), Rate of complications: 4.09 (±0.99)) were seen as most important (see Table 3). In contrast, general hospital information (Emergency services: 3.51 (±1.20), Hospital type: 3.45 (±1.12)), and narrative comments (3.56 (±1.12)) were rated less important. Not surprising, respondents who assigned highest importance to infection rates and technical

Table 2: Overview of the Selection of Hospitals (in %) (N = 2,700 Experiments, Equals 1,350 Finisher) (p value was Calculated Using Kruskal–Wallis)

Which Hospitals Did the Respondents Select?	*Quantitatively Dominant Hospital*	*p*	*Quantitatively Dominated Hospital*	*p*
Report card design: overall result				
Text information (N = 902)	48.1	.000	7.1	.072
Star display (N = 867)	47.3		8.5	
Numerical information (N = 931)	56.0		10.1	
Report cards without narrative comments				
Design 1: Text information without narrative comments (N = 443)	47.6	.270	6.8	.045
Design 3: Stars display without narrative comments (N = 444)	48.6		7.9	
Design 5: Numerical information without narrative comments (N = 463)	52.7		11.2	
Report cards with narrative comments				
Design 2: Text information with narrative comments (N = 459)	48.6	.000	7.4	.572
Design 4: Stars display with narrative comments (N = 423)	45.9		9.2	
Design 6: Numerical information with narrative comments (N = 468)	59.2		9.0	

quality of care information were more likely to have selected the quantitatively dominant hospital, whereas respondents who paid more attention to general hospital characteristics were more likely to have selected the quantitatively dominated hospital.

The Impact of Incorporating Narrative Comments on the Choice Behavior

There was no significant difference in the results for selecting the dominating hospital between report cards without narrative comments (49.7 percent; 671 of 1,350) and those displaying narratives (51.4 percent; 694 out of 1,350)

Table 3: Importance of Different Information Items for the Hospital Decision (Rated on a 1–5 Scale with 1 Not All Important and 5 Extremely Important) (Mean and SD; $N = 2{,}700$ Experiments, Equals 1,350 Finisher) (p value was Calculated Using Mann–Whitney U test)

	Overall				*Textual Design*				*Star Rating Design*				*Numerical Design*			
	Total	*Quantitatively Dominant Hospital*	*Quantitatively Dominated Hospital*	*p*	*Total*	*Quantitatively Dominant Hospital*	*Quantitatively Dominated Hospital*	*p*	*Total*	*Quantitatively Dominant Hospital*	*Quantitatively Dominated Hospital*	*p*	*Total*	*Quantitatively Dominant Hospital*	*Quantitatively Dominated Hospital*	*p*
General information—Emergency services	3.51 (1.20)	3.41 (1.16)	3.71 (1.13)	**	3.47 (1.19)	3.25 (1.21)	3.62 (1.10)	*	3.56 (1.21)	3.53 (1.15)	3.69 (1.17)		3.50 (1.18)	3.42 (1.12)	3.90 (1.03)	*
General information—Hospital type	3.45 (1.12)	3.33 (1.07)	3.71 (1.11)	**	3.41 (1.11)	3.22 (1.09)	3.62 (1.07)	*	3.53 (1.09)	3.47 (1.05)	3.64 (1.09)		3.42 (1.14)	3.30 (1.07)	3.86 (1.10)	**
Survey of patients' experience—Percentage of 9 or 10 ratings	3.82 (1.01)	3.95 (0.88)	3.86 (1.01)		3.82 (0.97)	3.95 (0.87)	3.71 (1.12)		3.84 (1.04)	3.94 (0.96)	4.00 (0.86)		3.80 (1.02)	3.88 (0.89)	4.07 (0.97)	
Survey of patients' experience—Percent who would definitely recommend the hospital	3.82 (1.03)	3.97 (0.90)	3.83 (0.96)		3.81 (1.00)	3.99 (0.88)	3.68 (1.09)		3.84 (1.04)	3.98 (0.93)	4.05 (0.76)		3.82 (1.04)	3.92 (0.94)	3.95 (0.99)	
Quality of care—Rate of unplanned readmission	3.98 (1.00)	4.19 (0.83)	3.99 (0.98)	*	3.97 (0.98)	4.27 (0.78)	3.85 (1.05)	*	3.95 (1.05)	4.09 (0.88)	4.23 (0.81)		4.02 (0.98)	4.17 (0.84)	4.17 (0.88)	*
Quality of care—Rate of complications	4.09 (0.99)	4.30 (0.81)	4.08 (0.93)	*	4.13 (0.95)	4.40 (0.72)	4.06 (0.95)	*	4.07 (1.02)	4.23 (0.87)	4.23 (0.84)		4.06 (1.10)	4.25 (0.85)	4.10 (0.96)	*
Costs of care—Provider in-network	3.91 (1.07)	3.98 (0.98)	3.98 (1.02)		3.91 (1.01)	4.00 (0.90)	3.71 (1.06)		3.95 (1.09)	3.96 (1.01)	4.33 (0.84)		3.88 (1.10)	3.95 (1.03)	4.00 (1.08)	
Costs of care—Out-of-pocket costs	3.95 (1.07)	4.01 (0.98)	3.99 (1.03)		3.99 (1.03)	4.04 (0.92)	3.71 (1.00)		3.98 (1.11)	4.00 (1.03)	4.36 (0.90)		3.89 (1.08)	3.94 (0.99)	3.90 (1.12)	

Continued

Table 3: *Continued*

	Overall				*Textual Design*				*Star Rating Design*				*Numerical Design*			
	Total	*Quantitatively Dominant Hospital*	*Quantitatively Dominated Hospital*	*p*	*Total*	*Quantitatively Dominant Hospital*	*Quantitatively Dominated Hospital*	*p*	*Total*	*Quantitatively Dominant Hospital*	*Quantitatively Dominated Hospital*	*p*	*Total*	*Quantitatively Dominant Hospital*	*Quantitatively Dominated Hospital*	*p*
Health care associated infections—Surgical site infection	4.12 (0.99)	4.30 (0.80)	4.05 (0.97)	**	4.19 (0.94)	4.43 (0.69)	3.74 (1.19)	*	4.04 (1.02)	4.14 (0.87)	4.46 (0.72)		4.12 (0.99)	4.29 (0.84)	4.14 (0.87)	**
Health care associated infections—Intestinal infections	4.09 (1.01)	4.26 (0.85)	4.09 (1.00)	*	4.16 (0.93)	4.37 (0.74)	3.94 (1.01)	*	4.02 (1.06)	4.08 (0.96)	4.38 (0.91)		4.10 (1.02)	4.25 (0.91)	4.26 (0.86)	
Patient stories—Narrative Comments	3.56 (1.12)	3.59 (1.04)	3.55 (1.12)		3.51 (1.12)	3.48 (1.07)	3.29 (1.22)		3.62 (1.12)	3.73 (0.98)	3.69 (1.06)		3.56 (1.12)	3.56 (1.06)	3.74 (1.10)	

Note: Importance of items is rated on a 1–5 scale [1 not all important; 5 extremely important].
*$p < .05$, **$p < .001$.

($p = .376$). But there was a statistically significant interaction between exposure to narratives and the format for presenting clinical performance (Table 2). When report cards do not incorporate narratives, there are moderate (5 percentage point) differences across design in selection of both dominant and dominated hospitals, though only for the latter are these cross-design differences statistically significant. By contrast, when narratives are included, the cross-design differences for dominated hospitals disappear, but those for selecting the dominant hospital become much larger in magnitude, with the numeric design having almost a 15 percentage point difference (compared to star ratings) in selection of the hospital with the best clinical ratings.

These same results can be viewed from an alternative perspective by comparing how hospital choice varies *within* each report card design, when narratives are introduced. The likelihood for selecting the hospital with the highest quantifiable metrics significantly increase when comments were combined with numerical information (59.2 percent vs. 52.7 percent, $p = .046$), but not when combined with textual information (48.6 percent vs. 47.6 percent, $p = .774$) or a star rating display (45.9 percent vs. 48.6 percent, $p = .412$).

Exploring/Explaining Cross-Arm Experimental Differences

The logistic regression models estimated to explore the underpinnings of these experimental effects (see Table 4). Accounting for the impact of demographics, health experiences and decision-styles did little to alter the baseline results: narratives had relatively little impact on hospital choice—the slightly higher (selecting the quantitatively dominant hospital) or lower (selecting the dominated) odds for those report cards containing narrative comments were not proven to be statistically significant in any model ($p > .05$ for each). By contrast, the impact of presentation format was consistent and significant—but complicated. Even after accounting for participant's attributes, experiences, and decision styles, those presented with numerical performance metrics were between 1.37 (95 percent CI: 1.14–1.65, $p < .001$) and 1.43 (95 percent CI: 1.18–1.73, $p < .001$) times more likely to select the dominant hospital than for those with textual quality information (the comparison group in the model); even larger differences were evident compared to the star-rated report designs. But this format was also associated with significantly higher odds of selecting the quantitatively dominated hospital: here, the odds ranged between 1.41 (95 percent CI: 1.01–1.98, $p < .05$) and 1.47 (95 percent CI: 1.06–2.05, $p < .05$).

Table 4: Logistic Regression Models Predicting Likelihood of Selection of the Quantitatively Dominant Hospital and Quantitatively Dominated Hospital Based on the Implementation of Narrative Comments and the Different Design Types (N = 2,700)

	Choice of the Quantitatively Dominant Hospital				*Choice of the Quantitatively Dominated Hospital*			
	Model 1 OR (95% CI)	*Model 2*[†] OR (95% CI)	*Model 3*[‡] OR (95% CI)	*Model 4*[§] OR (95% CI)	*Model 1 OR (95% CI)*	*Model 2*[†] OR (95% CI)	*Model 3*[‡] OR (95% CI)	*Model 4*[§] OR (95% CI)
Narrative comments								
Narrative comments not displayed (N = 1,350)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Narrative comments displayed (N = 1,350)	1.07 (0.92;1.24)	1.07 (0.92;1.25)	1.06 (0.91;1.25)	1.07 (0.91;1.25)	0.98 (0.75;1.29)	0.99 (0.75;1.30)	0.99 (0.75;1.30)	0.99 (0.75;1.30)
Design types								
Textual information (N = 902)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)	1.00 (ref)
Star ratings (N = 867)	0.97 (0.80;1.17)	0.96 (0.79;1.16)	0.97 (0.80;1.18)	0.95 (0.78;1.16)	1.22 (0.86;1.73)	1.24 (0.87;1.77)	1.19 (0.83;1.70)	1.19 (0.83;1.71)
Numerical information (N = 931)	1.37 (1.14;1.65)*	1.42 (1.17;1.71)**	1.43 (1.18;1.73)**	1.42 (1.17;1.73)**	1.47 (1.06;2.05)*	1.45 (1.04;2.03)*	1.41 (1.01;1.98)*	1.41 (1.01;1.98)*
Statistical model details	$\chi^2(3) = 17.514$ $R^2 = .009$; $p = .001$	$\chi^2(21) = 169.367$ $R^2 = .081$; $p = .000$	$\chi^2(26) = 186.566$ $R^2 = .091$; $p = .000$	$\chi^2(31) = 282.621$, $R^2 = .135$; $p = .000$	$\chi^2(3) = 5.294$ $R^2 = .004$; $p = .151$	$\chi^2(21) = 51.712$ $R^2 = .043$; $p = .000$	$\chi^2(26) = 65.563$ $R^2 = .055$; $p = .000$	$\chi^2(31) = 70.563$ $R^2 = .060$; $p = .000$

[†]Model 2: Adjusted for demographics (age, gender, marital status, education, ethnicity, health insurance).
[‡]Model 3: Adjusted for demographics, health-related demographics (chronic conditions and number of hospital treatments in the previous 3 years).
[§]Model 4: Adjusted for demographics, health-related demographics, high versus low maximizer and perceived differences in the quality of care.
*$p < .05$, **$p < .001$.

Table 5: Descriptive Analysis for Choosing the Dominating Hospital and Multivariate Logistic Regression Analyses; Association between the Selection of the Dominating Hospital and Education

	Mean (in percent)	*Model 1*† OR (95% CI)	*Model 2*‡ OR (95% CI)	*Model 3*§ OR (95% CI)
Design 1: Text information without narrative comments ($N = 443$)				
Education	*	*	*	
Less than high school	35.8	1.00 (ref)	1.00 (ref)	1.00 (ref)
High school graduate	39.4	1.26 (0.65;2.47)	1.24 (0.62;2.48)	1.13 (0.55;2.32)
Some college or associate's degree	51.2	2.06 (1.04;4.07)*	1.91 (0.94;3.89)	1.52 (0.73;3.18)
Completed college/ advanced degree	60.3	2.82 (1.37;5.83)*	2.70 (1.28;5.71)*	2.28 (1.04;4.98)*
Design 2: Text information with narrative comments ($N = 459$)				
Education	**	**	**	**
Less than high school	23.3	1.00 (ref)	1.00 (ref)	1.00 (ref)
High school graduate	37.7	1.92 (0.98;3.78)	1.84 (0.93;3.65)	1.81 (0.90;3.66)
Some college or associate's degree	63.9	6.24 (3.06;12.70)**	6.34 (3.09;12.99)**	5.35 (2.56;11.21)**
Completed college/ advanced degree	60.0	4.77 (2.26;10.09)**	4.66 (2.19;9.94)**	4.47 (2.04;9.80)
Design 3: Stars display without narrative comments ($N = 444$)				
Education	**	*	*	*
Less than high school	26.9	1.00 (ref)	1.00 (ref)	1.00 (ref)
High school graduate	49.3	2.52 (1.30;4.87)*	2.53 (1.29;4.96)*	2.34 (1.17;4.68)*
Some college or associate's degree	57.1	2.84 (1.41;5.72)*	3.06 (1.49;6.30)*	2.88 (1.36;6.10)*
Completed college/ advanced degree	51.9	2.07 (0.99;4.32)	2.22 (1.05;4.70)*	2.10 (0.97;4.56)
Design 4: Stars display with narrative comments ($N = 423$)				
Education				
Less than high school	44.6	1.00 (ref)	1.00 (ref)	1.00 (ref)
High school graduate	42.4	0.92 (0.49;1.74)	0.87 (0.45;1.70)	0.82 (0.41;1.63)
Some college or associate's degree	43.0	0.86 (0.45;1.68)	0.73 (0.36;1.46)	0.65 (0.32;1.33)
Completed college/ advanced degree	54.3	1.26 (0.62;2.59)	1.27 (0.60;2.69)	1.11 (0.51;2.42)

Continued

Table 5: *Continued*

	Mean (in percent)	*Model 1*† OR (95% CI)	*Model 2*‡ OR (95% CI)	*Model 3*§ OR (95% CI)
Design 5: Numerical information without narrative comments ($N = 463$)				
Education	**	*	*	*
Less than high school	40.3	1.00 (ref)	1.00 (ref)	1.00 (ref)
High school graduate	42.5	1.09 (0.61;1.94)	1.04 (0.57;1.89)	1.06 (0.57;1.95)
Some college or associate's degree	62.4	2.50 (1.34;4.66)*	2.42 (1.27;4.58)*	2.16 (1.11;4.17)*
Completed college/ advanced degree	64.8	2.58 (1.28;5.18)*	2.41 (1.19;4.90)*	2.17 (1.04;4.54)*
Design 6: Numerical information with narrative comments ($N = 468$)				
Education	*		*	
Less than high school	41.1	1.00 (ref)	1.00 (ref)	1.00 (ref)
High school graduate	56.4	1.80 (0.97;3.33)	2.08 (1.11;3.90)*	1.43 (0.73;2.79)
Some college or associate's degree	64.6	2.26 (1.16;4.40)*	2.83 (1.41;5.68)*	1.82 (0.86;3.83)
Completed colelge/ advanced degree	67.7	2.18 (1.10;4.33)*	2.57 (1.27;5.21)*	1.68 (0.80;3.53)

†Model 1: Adjusted for demographics (age, gender, marital status, education, ethnicity, health insurance).
‡Model 2: Adjusted for demographics, health-related demographics (chronic conditions and number of hospital treatments in the previous 3 years).
§Model 3: Adjusted for demographics, health-related demographics, high versus low maximizer, and perceived differences in the quality of care.
*$p < .05$, **$p < .001$.

Interaction of Education, Narratives, and Report Card Design

Because the educational gradient was so-pronounced in our baseline analyses, we explored the interaction of report card attributes and this educational gradient in a set of additional regression models, with the sample stratified by experimental arm (see Tables 2.1–2.6 for the complete model results). The key findings for education are presented in Table 5. Higher educational levels were associated with hospital choices more responsive to performance metrics, especially having at least some college or associate's degree. But the choices of the least educated respondents—and therefore the educational gradients in choice—appear to be very sensitive to the intersection of report card design and narrative commentary. Including narratives seems to erode

decision quality among the least educated respondents when performance metrics are in narrative form, but enhance it for star ratings. By contrast, when performance metrics are presented in numeric formats, their impact on choice appears unaltered by the inclusion of narratives.

DISCUSSION

Our results confirm previous findings showing that consumers often do not select the highest rated health care provider when presented with choices sets that incorporate options that are dominant on the quantitative metrics (Damberg and McNamara 2014; Kanouse et al. 2016). For this experiment, in slightly more than half of all cases (50.6 percent), respondents selected the hospital with the best clinical results. Across countries, similar experimental studies (Hibbard et al. 2001; Hibbard, Stockard, and Tusler 2005; Gerteis et al. 2007; Peters et al. 2007; Hibbard, Greene, and Daniel 2010; Damman et al. 2011; Emmert et al. 2014b; Schlesinger et al. 2014) have shown roughly comparable patterns. In our study, the frequency with which the dominant hospital in terms of clinical quality was not selected might reflect in part the complexity of the choice task: respondents had to compare six attributes across hospitals.[1]

How closely any experimental findings comport with real-world choices is difficult to assess. Certainly there is ample evidence from other health-related choices (e.g., Part D drug coverage under Medicare) that suboptimal choices are quite common (Schlesinger et al. 2014). But evidence on real-world choices among hospitals is quite limited. The platforms supporting choice among hospitals (e.g., Hospital Compare) are often far *more* complex than our experiment and consumers are presented with a far larger array of options—one might expect this to make choices even more daunting for less sophisticated (including less educated) consumers. On the other hand, those websites often allow consumers to limit their comparisons to a subset of providers and performance metrics. This simplifies choice, though this sort of "filtering" may yield its own limitations and biases (Schlesinger et al. 2014).

As suggested above, the design of the report cards seems to influence on the comprehensibility of the quality information. Those report cards with a numerical design appeared to be more comprehensible (56.0 percent) than those with textual information (48.1 percent) or a star display (47.3 percent). This result is in line with one recent study that found words and numerical

information to be slightly more comprehensible than stars (Gerteis et al. 2007). But the broader evidence from the field is mixed.

A qualitative study by Mazor and Dodd (2009) also concluded that numeric presentations are preferable to symbols since some consumers perceive symbols to be confusing or difficult to interpret; some even wondered whether symbols were used to mask information that might otherwise impugn providers' reputations. However, other studies found numerical information to be less comprehensible (Hibbard and Sofaer 2010; Donelan et al. 2011). It is difficult to discern the causes of these different findings, because the studies used different presentation formats or study designs (Peters et al. 2007; Damman et al. 2011; Emmert et al. 2014b).

We did not identify any major effects from incorporating narratives into hospital report cards—suggesting that the current proliferation of patient comments does not pose a major barrier to consumers' selecting hospitals with the highest clinical ratings. However, this finding may also be caused, at least to some extent, by our study approach. Because we displayed two positive and one negative comments for each hospital, respondents might have concluded that consumers leaving comments were not very discerning about quality differences among hospitals and therefore conclude that there was not much added value in this information, choosing therefore to focus more on technical quality of care information or infection rates when selecting a hospital. These clinical metrics might also have seemed more important for choice of hospital than when selecting a physician (Kanouse et al. 2016), because the threats of poor technical quality were more pronounced for inpatient care.

Our findings suggest that with some report card designs, the inclusion of comments increases consumer's selection of hospitals with high clinical ratings. This contrasts with studies of clinician choice (Schlesinger et al. 2014; Kanouse et al. 2016), where comments have been documented to divert consumers' attention away from quantified performance metrics. That could reflect differences in experimental design—or differences in the types of providers being chose, since consumer comments about clinicians touch on dimensions (e.g., empathy, interactional style) that may be seen by consumers as more essential than the hospital attributes characterized in most comments about inpatient care.

Nevertheless, the impact of narratives was more significant, when interacted with particular report card designs. Adding narratives induced a significant increase in selection of the top-rated clinical hospital, when performance metrics were presented as numerical information, but not when combined with textual information or a star rating display. In other words, the magnitude

of the design effects becomes far more pronounced in a context where narratives are also available. This suggests that report card designers might need to revisit the question of display format, as patient comments become ubiquitous over the internet.

As in previous research, we found that consumers' education affected their interpretation of the information presented on report cards (Hibbard, Stockard, and Tusler 2005; Damman et al. 2011). More educated consumers were more likely to select hospitals with the highest clinical ratings. This reflects in part that more educated consumers see more value in these performance metrics (see Table S3): for our respondents, the average value placed on these quality metrics was about 10 percent greater for the most educated respondents compared to the least. But since the magnitude of the educational gradient varies across both report card design and incorporation of narratives, these patterns in choice cannot simply reflect differences in preferences by education. It remains unclear why these educational gradients vary, suggesting the need for additional research about how consumers interpret performance metrics when they are combined with narratives.

Our findings should be considered in light of some methodological limitations. First, our study was designed as an online survey, so that the results might be influenced by self-selection of the study population. Even though the surveyed online panel was recruited through several different recruitment strategies some groups in the population may still be under-represented (e.g., the elderly, the less educated). Second, even though our surveyed sample does not statistically significant differ to the American public in terms of age, gender, education, and marital status, there were significant differences with respect to ethnicity and household size. We controlled for these in our regression models, but they may have influenced our findings in more subtle ways.

Third, our study was designed as a cross-sectional survey. Thus, we were able to identify associations between exposure and outcomes but not infer cause and effect. Fourth, although our experimental report card was designed to mimic Hospital Compare, it differs in the scope and presentation of quality information, making generalizations to any real-world site more speculative. Fifth, we created standardized narrative comments based on real-world comments to control for the heterogeneity of narrative comments. Even though this enabled us to detect and compare the impact of the comments for different presentation designs, their impact in the real world might be different since narratives are more unstructured (Greaves, Millett, and Nuki 2014) and cover a broader range of topics. In addition, assigning comments to hospitals in ways

that correlated either positively or negatively with clinical quality metrics might also have led consumers to see comments as more salient, giving them more impact on choice (Kanouse et al. 2015). Finally, due to multiple tests, some of our findings should be considered exploratory and confirmed before presuming these interactions are replicable.

Conclusion

As recently stated by Greaves "[. . .] the jury is still out on where narrative comments fit in the complex landscape of quality measurement" (Greaves, Millett, and Nuki 2014). Our findings suggest that this answer will depend on how other performance metrics are formatted and whose choices are being examined. To be able to better understand the impact of narratives, researchers will need to learn more about the association between the narrative ratings and clinical quality of care measures.

Consumers clearly struggle when making complex health care decisions, even in a simplified experimental environment. Only half of the respondents in this study selected the hospital with the best clinical results and every twelfth respondent selected the quantitatively dominated hospital (whether or not narratives were present). The design of the report cards has relevance to presenting quality information in a comprehensible way to the public, even though the differences across formats identified here ranged from small (5 percentage points) to moderate (15 percentage points) in magnitude.

Although narratives do not seem to represent a "proximate threat" to other performance metrics in consumers' choices among hospitals, they do appear to exert more subtle influences, particularly for certain report card designs and particular subsets of consumers. Better understanding these effects will be essential, since patient comments are almost certain to proliferate over the internet (Schlesinger et al. 2015). How best to respond will be a question that policymakers, report card designers, quality alliances, and insurers will all need to come to terms within the coming years.

Acknowledgments

Joint Acknowledgment/Disclosure Statement: Martin Emmert reports receiving support from a 2014/2015 Commonwealth Fund/Robert Bosch Foundation

Harkness Fellowship in Health Care Policy and Practice. The views presented here are those of the authors and should not be attributed to The Commonwealth Fund or its directors, officers, or staff.

Disclosures: None.

Disclaimers: None.

NOTE

1. By comparison, Gerteis et al. (2007) determined a comprehensibility range from 47 to 89 percent when five quality indicators for 10 nursing homes were presented; however, no information regarding either patient satisfaction or costs of care was included.

REFERENCES

Archak, N., A. Ghose, and P. G. Ipeirotis. 2011. "Deriving the Pricing Power of Product Features by Mining Consumer Reviews." *Management Science* 57 (8): 1485–509.

Brennan, P. F. 1995. "Patient Satisfaction and Normative Decision Theory." *Journal of the American Medical Informatics Association* 2 (4): 250–9.

Brutus, S. 2010. "Words versus Numbers: A Theoretical Exploration of Giving and Receiving Narrative Comments in Performance Appraisal." *Human Resource Management Review* 20 (2): 144–57.

Center for Advancing Health. 2009. *Getting Tools Used: Lessons for Health Care from Successful Consumer Decision Aids*. Washington, D.C.: Center for Advancing Health.

Damberg, C. L., and P. McNamara. 2014. "Postscript: Research Agenda to Guide the Next Generation of Public Reports for Consumers." *Medical Care Research and Review* 71 (5 Suppl.): 97S–107S.

Damman, O. C., M. Hendriks, J. Rademakers, P. Spreeuwenberg, D. M. Delnoij, and P. P. Groenewegen. 2011. "Consumers' Interpretation and Use of Comparative Information on the Quality of Health Care: The Effect of Presentation Approaches." *Health Expectations: An International Journal of Public Participation in Health Care and Health Policy* 15 (2): 197–211.

Detz, A., A. López, and U. Sarkar. 2013. "Long-Term Doctor-Patient Relationships: Patient Perspective from Online Reviews." *Journal of Medical Internet Research* 15 (7): e131.

Dieckmann, N. F., P. Slovic, and E. M. Peters. 2009. "The Use of Narrative Evidence and Explicit Likelihood by Decisionmakers Varying in Numeracy." *Risk Analysis* 29: 1473–88.

Donelan, K., R. S. Rogers, A. Eisenhauer, E. Mort, and A. K. Agnihotri. 2011. "Consumer Comprehension of Surgeon Performance Data for Coronary Bypass Procedures." *Annals of Thoracic Surgery* 91 (5): 1400–6.

Emmert, M., F. Meier, A. K. Heider, C. Dürr, and U. Sander. 2014a. "What Do Patients Say about Their Physicians? An Analysis of 3000 Narrative Comments Posted on a German Physician Rating Website." *Health Policy* 118 (1): 66–73.

Emmert, M., S. Hessemer, N. Meszmer, and U. Sander. 2014b. "Do German Hospital Report Cards Have the Potential to Improve the Quality of Care?" *Health Policy* 118 (3): 386–95.

Faber, M., M. Bosch, H. Wollersheim, S. Leatherman, and R. Grol. 2009. "Public Reporting in Health Care: How Do Consumers Use Quality-of-Care Information? A Systematic Review." *Medical Care* 47 (1): 1–9.

Fox, S. 2011. *The Social Life of Health Information.* Washington, DC: Pew Research Center's Internet & American Life Project.

Gerteis, M., J. S. Gerteis, D. Newman, and C. Koepke. 2007. "Testing Consumers' Comprehension of Quality Measures Using Alternative Reporting Formats." *Health Care Financing Review* 28 (3): 31–45.

Greaves, F., C. Millett, and P. Nuki. 2014. "England's Experience Incorporating "Anecdotal" Reports from Consumers into Their National Reporting System: Lessons for the United States of What to Do or Not to Do?" *Medical Care Research and Review* 71 (5 Suppl.): 65S–80S.

Hanoch, Y., T. Rice, J. Cummings, and S. Wood. 2009. "How Much Choice Is Too Much? The Case of the Medicare Prescription Drug Benefit." *Health Services Research* 44 (4): 1157–68.

Hibbard, J. H., J. Greene, and D. Daniel. 2010. "What Is Quality Anyway? Performance Reports That Clearly Communicate to Consumers the Meaning of Quality of Care." *Medical Care Research and Review* 67 (3): 275–93.

Hibbard, J. H., and E. Peters. 2003. "Supporting Informed Consumer Health Care Decisions: Data Presentation Approaches That Facilitate the Use of Information in Choice." *Annual Review of Public Health* 24: 413–33.

Hibbard, J. H., and S. Sofaer. 2010. *Best Practices in Public Reporting No. 1: How to Effectively Present Health Care Performance Data to Consumers.* AHRQ Publications No. 10-0082-EF. Rockville, MD: Agency for Healthcare Research and Quality.

Hibbard, J. H., J. Stockard, and M. Tusler. 2005. "It Isn't Just about Choice: The Potential of a Public Performance Report to Affect the Public Image of Hospitals." *Medical Care Research and Review* 62 (3): 358–71.

Hibbard, J. H., P. Slovic, E. Peters, M. L. Finucane, and M. Tusler. 2001. "Is the Informed-Choice Policy Approach Appropriate for Medicare Beneficiaries?" *Health Affairs (Project Hope)* 20 (3): 199–203.

Hibbard, J. H., J. Greene, S. Sofaer, K. Firminger, and J. Hirsh. 2012. "An Experiment Shows That a Well-Designed Report on Costs and Quality Can Help Consumers Choose High-Value Health Care." *Health Affairs (Project Hope)* 31 (3): 560–8.

Kanouse, D. C., M. Schlesinger, D. Shaller, S. C. Martino, and L. Rybowski. 2016. "How Patient Comments Affect Consumers' Use of Physician Performance Measures." *Medical Care* 54 (1): 24–31.

King, R. A., P. Racherla, and V. D. Bush. 2014. "What We Know and Don't Know about Online Word-of-Mouth: A Review and Synthesis of the Literature." *Journal of Interactive Marketing* 28 (3): 167–83.

Lagu, T., and P. K. Lindenauer. 2010. "Putting the Public Back in Public Reporting of Health Care Quality." *Journal of the American Medical Association* 304 (15): 1711–2.

Lagu, T., N. S. Hannon, M. B. Rothberg, and P. K. Lindenauer. 2010. "Patients' Evaluations of Health Care Providers in the Era of Social Networking: An Analysis of Physician-Rating Websites." *Journal of General Internal Medicine* 25 (9): 942–6.

Lagu, T., S. L. Goff, N. S. Hannon, A. Shatz, and P. K. Lindenauer. 2013. "A Mixed-Methods Analysis of Patient Reviews of Hospital Care in England: Implications for Public Reporting of Health Care Quality Data in the US." *Joint Commission Journal on Quality and Patient Safety* 39 (1): 7–15.

Martino, S. C. 2012. *Consumers' Appraisal of Anecdotal Accounts of Patient Experience.* Bethesda, MD: AHRQ Annual Conference.

Mazor, K. M., and K. S. Dodd. 2009. "A Qualitative Study of Consumers' Views on Public Reporting of Health Care-Associated Infections." *American Journal of Medical Quality* 24 (5): 412–8.

Mehrotra, A., P. S. Hussey, A. Milstein, and J. H. Hibbard. 2012. "Consumers' and Providers' Responses to Public Cost Reports, and How to Raise the Likelihood of Achieving Desired Results." *Health Affairs (Project Hope)* 31 (4): 843–51.

Peters, E., N. Dieckmann, A. Dixon, J. H. Hibbard, and C. K. Mertz. 2007. "Less Is More in Presenting Quality Information to Consumers." *Medical Care Research and Review* 64 (2): 169–90.

Schlesinger, M., D. E. Kanouse, L. Rybowski, S. C. Martino, and D. Shaller. 2012. "Consumer Response to Patient Experience Measures in Complex Information Environments." *Medical Care* 50 (Suppl.): S56–64.

Schlesinger, M., D. E. Kanouse, S. C. Martino, D. Shaller, and L. Rybowski. 2014. "Complexity, Public Reporting, and Choice of Doctors: A Look inside the Blackest Box of Consumer Behavior." *Medical Care Research and Review* 71 (5 Suppl.): 38S–64S.

Schlesinger, M., R. Grob, D. Shaller, S. C. Martino, A. M. Parker, M. L. Finucane, J. L. Cerully, and L. Rybowski. 2015. "Taking Patients' Narratives about Clinicians from Anecdote to Science." *New England Journal of Medicine* 373 (7): 675–9.

Shaffer, V. A., J. Owens, and B. J. Zikmund-Fisher. 2013. "The Effect of Patient Narratives on Information Search in a Web-Based Breast Cancer Decision Aid: An Eye-Tracking Study." *Journal of Medical Internet Research* 15 (12): e273.

Slovic, P. 1982. "Towards Understanding and Improving Decisions." In *Human Performance and Productivity. Information Processing and Decision Making*, edited by W. C. Howell and E. A. Fleishman, pp. 157–83. Hillsdale, NJ: Erlbaum.

U.S. Census Bureau. 2012. *Current Population Survey.* Washington, DC. Available at https://www.census.gov/population/age/data/cps.html

Winterbottom, A., H. L. Bekker, M. Conner, and A. Mooney. 2008. "Does Narrative Information Bias Individual's Decision Making? A Systematic Review." *Social Science & Medicine* 67 (12): 2079–88.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Table S1: Characteristics of Respondents According to Their Hospital Choice (p value was calculated using chi-square test and Kruskal–Wallis) ($N = 1{,}350$).

Table S2.1: Multivariate Regression Analyses—Design 1; Adjusted Odds Ratio (OR), 95% Confidence Interval (CI), and p-value of the Association between the Selection of the Quantitatively Dominant Hospital and Dependent Variables [if no p-value is presented it means that no statistically significant differences could be detected].

Table S2.2: Multivariate Regression Analyses—Design 2; Adjusted Odds Ratio (OR), 95% Confidence Interval (CI), and p-value of the Association between the Selection of the Quantitatively Dominant Hospital and Dependent Variables [if no p-value is presented it means that no statistically significant differences could be detected].

Table S2.3: Multivariate Regression Analyses—Design 3; Adjusted Odds Ratio (OR), 95% Confidence Interval (CI), and p-value of the Association between the Selection of the Quantitatively Dominant Hospital and Dependent Variables [if no p-value is presented it means that no statistically significant differences could be detected].

Table S2.4: Multivariate Regression Analyses—Design 4; Adjusted Odds Ratio (OR), 95% Confidence Interval (CI), and p-value of the Association between the Selection of the Quantitatively Dominant Hospital and Dependent Variables [if no p-value is presented it means that no statistically significant differences could be detected].

Table S2.5: Multivariate Regression Analyses—Design 5; Adjusted Odds Ratio (OR), 95% Confidence Interval (CI), and p-value of the Association between the Selection of the Quantitatively Dominant Hospital and Dependent Variables [if no p-value is presented it means that no statistically significant differences could be detected].

Table S2.6: Multivariate Regression Analyses—Design 6; Adjusted Odds Ratio (OR), 95% Confidence Interval (CI), and p-value of the Association between the Selection of the Quantitatively Dominant Hospital and Dependent Variables [if no p-value is presented it means that no statistically significant differences could be detected].

Table S3: Importance of Different Information Items for the Hospital Decision (rated on a 1–5 scale with 1 not all important and 5 extremely important) (Mean and SD; $N = 1{,}350$ respondents) (p value was calculated using using Kruskal–Wallis).